Genome-Wide Profiling of Pluripotent Cells Reveals a Unique Molecular Signature of Human Embryonic Germ Cells

Citation

Published Version
doi:10.1371/journal.pone.0039088

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:10436260

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Genome-Wide Profiling of Pluripotent Cells Reveals a Unique Molecular Signature of Human Embryonic Germ Cells

Nikta Pashai1, Haiping Hao2, Angelo All1,3, Siddharth Gupta1, Raghothama Chaerkady8, Alejandro De Los Angeles9,10,11, John D. Gearhart5,6, Candace L. Kerr4,7*

1 Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 2 Deep Sequencing and Microarray Core, High Throughput Biology Center, Johns Hopkins University, Baltimore, Maryland, United States of America, 3 Department of Neurology, Johns Hopkins University, Baltimore, Maryland, United States of America, 4 Stem Cell Program, Institute for Cell Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America, 5 Department of Cell and Developmental Biology, Institute of Regenerative Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 6 Department of Animal Biology, Institute of Regenerative Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 7 Department of Gynecology and Obstetrics, Institute for Cell Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America, 8 Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 9 Stem Cell Transplantation Program, Division of Pediatric Hematology Oncology, Children’s Hospital Boston, Massachusetts, United States of America, 10 Department of Biological Chemistry and Molecular Pharmacology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, United States of America, 11 Harvard Stem Cell Institute, Cambridge, Massachusetts, United States of America

Abstract

Human embryonic germ cells (EGCs) provide a powerful model for identifying molecules involved in the pluripotent state when compared to their progenitors, primordial germ cells (PGCs), and other pluripotent stem cells. Microarray and Principal Component Analysis (PCA) reveals for the first time that human EGCs possess a transcription profile distinct from PGCs and other pluripotent stem cells. Validation with qRT-PCR confirms that human EGCs and PGCs express many pluripotency-associated genes but with quantifiable differences compared to pluripotent embryonic stem cells (ESCs), induced pluripotent stem cells (IPSCs), and embryonal carcinoma cells (ECCs). Analyses also identified a number of target genes that may be potentially associated with their unique pluripotent states. These include IPO7, MED7, RBM26, HSPD1, and KRAS which were upregulated in EGCs along with other pluripotent stem cells when compared to PGCs. Other potential target genes were also found which may contribute toward a primed ESC-like state. These genes were exclusively up-regulated in ESCs, IPSCs and ECCs including PARP1, CCNE1, CDK6, AURKA, MAD2L1, CCNG1, and CCNB1 which are involved in cell cycle regulation, cellular metabolism and DNA repair and replication. Gene classification analysis also confirmed that the distinguishing feature of EGCs compared to ESCs, ECCs, and IPSCs lies primarily in their genetic contribution to cellular metabolism, cell cycle, and cell adhesion. In contrast, several genes were found upregulated in PGCs which may help distinguish their unipotent state including HBA1, DMRT1, SPANXA1, and EHD2. Together, these findings provide the first glimpse into a unique genomic signature of human germ cells and pluripotent stem cells and provide genes potentially involved in defining different states of germ-line pluripotency.

Introduction

Primordial germ cells (PGCs) are unipotent progenitors of sperm and egg which retain an innate ability to generate pluripotent stem cells in vivo, called embryonal carcinomas (ECCs), and in vitro, known as embryonic germ cells (EGCs). It is unknown whether mechanisms similar to those involved in the generation of these cells are also involved in maintaining the pluripotent status of other stem cells such as embryonic stem cells (ESCs) and induced pluripotent stem cells (IPSCs). In this study, the first genome wide assessments of human PGCs and EGCs were performed and compared with other pluripotent ESCs, IPSCs and ECCs.

PGCs are unipotent in that they are lineage-restricted to become germ cells. They do not exhibit self-renewal and do not survive past one week under standard tissue culture conditions [1]. In the mouse, it is known that PGCs are derived from a region of the epiblast that mainly gives rise to the extra-embryonic mesoderm. In humans, PGCs first appear between the third and fourth week post-fertilization in the endoderm of the dorsal wall of the yolk sac, near the allantoid, then proceed to migrate through the hindgut during the fourth week and dorsal mesentery in the fifth week to reach the genital ridge [2–4].

* E-mail: ckerr@jhmi.edu
Under appropriate conditions, human PGCs can generate pluripotent stem cells. EGCs are pluripotent stem cells derived from PGCs in cell culture using specific growth factors and media formulations. The term EGCs, was given by Donovan et al. [5,6] who derived them from mouse PGCs, in order to distinguish them from ESCs. Since then, a handful of other laboratories have also reported the generation of human EGC lines [7–11]. EGC lines are another source of pluripotent cells that are cultured from adult teratocarcinomas from which there is genetic, immunological, and morphological evidence suggesting a PGC origin [12]. Lastly, IPSCs can be generated from PGCs by lentiviral transduction of pluripotent regulators similar to those used to generate IPSCs from somatic cells [13–18].

EGCs share the general properties of pluripotent stem cells including unlimited self-renewal and the ability to give rise to cells that represent all three embryonic germ layers. In contrast, EGCs are unlike any other pluripotent stem cell as they are derived from differentiated cells without IPSC-like targeted genetic manipulations and unlike ECCs, maintain a stable karyotype. Furthermore, while human EGCs can generate a variety of cell-types representative of the three germ layers in vitro, they do not form teratomas in vivo like their mouse EGC counterparts. Therefore, EGCs may exist in a uniquely, partial, or intermediate pluripotent state. As such, comparisons between EGCs and PGCs with other pluripotent stem cells provide a powerful model to identify factors that are associated with different states of pluripotency.

Distinct states of pluripotency have been revealed by several laboratories which have shown that pluripotent stem cells exhibit differences in their clonal or self-renewing and differentiating capacities [19–22]. For instance, mouse ESCs and IPSCs in the “naïve state” demonstrate single cell clonal ability, rounded colony capacities [19–22]. For instance, mouse ESCs and IPSCs in the differences in their clonal or self-renewing and differentiating laboratories which have shown that pluripotent stem cells exhibit differences and that represent all three embryonic germ layers in vitro. They do not form teratomas in vivo like their mouse EGC counterparts. Therefore, EGCs may exist in a uniquely, partial, or intermediate pluripotent state. As such, comparisons between EGCs and PGCs with other pluripotent stem cells provide a powerful model to identify factors that are associated with different states of pluripotency.

In contrast, conventional human ESCs and IPSCs and mouse epiblast-derived stem cells exist in a “primed state” of pluripotency exhibiting flattened colony morphology, insufficient clonal expansion, and a dependence on FGF2 and TGF/Activin signaling. These differences in pluripotent states have been attributed to species differences as well as the developmental state of the stem cell origin and yet they are inter-convertible depending on the cell culture environment. For instance, the primed state of human ESCs and IPSCs was shown to be convertable to the naïve mouse ESC-like state given the appropriate factors in cell culture [22]. It has also been shown that mouse EGCs will behave similar to the naïve state of mouse ESCs under similar culture conditions [23]. However, it remains unknown whether human EGCs could also be converted to a naïve state. Indeed, there is considerable interest in deciphering the range of multiple pluripotent states in human cells as they could be utilized to partition out mechanisms that regulate distinct attributes of the pluripotent phenotype.

Currently, the pluripotent state of conventional human EGCs is unknown. For instance, like human ESCs, conventional human EGCs express SSEA3, SSEA4 and TRA antigens, TRA-1-60 and TRA-1–80, which are inefficient at clonal expansion and require FGF2 in cell culture [8,24]. However, similar to mouse ESCs, human EGCs share rounded morphology, express SSEA1 and require LIF for their survival. Given that human EGCs share features in common with both mouse ESCs and human ESCs, it is likely that conventional EGCs fall in their own unique state of pluripotency. Therefore, the following study provides new insight into this question and reveals the genomic signature of EGCs which will identify new candidate genes for regulating pluripotency.

Comparisons between EGCs and PGCs will also help establish a unique signature of human PGCs which have not been demonstrated before while also providing further insight into whether ESCs originate from a PGC origin. Indeed, several lines of evidence suggest that EGCs and ESCs may originate from an early germ cell progenitor [23–27]. For instance, several reports have demonstrated that mouse ESCs express genes associated with immature male and female germ cells such as Stella, deleted in azoospermia-like (Dazl) and Frags [28,29]. Similarly, one study that examined the differentiation of human ESCs into germ cells [30], detected the expression of eight genes characteristic of early germ cells in ESCs and none from six genes expressed by later germ cells. Most significantly, this study demonstrated gene expression of Dazl by ESCs but not by human inner cell mass (ICM). In a recent study, Scholer et al. [31] demonstrated significant similarities in gene profiles between mouse PGCs and ESCs. Despite these relevant studies, it is largely unknown how human PGCs resemble ESCs at the molecular level compared to other PGC-derived stem cells.

Global expression analyses have been extensively performed to examine the molecular signature of human pluripotent stem cells. These studies have characterized gene expression profiles of human ESC, IPSC, and ECC lines using microarray and other high-throughput analyses in attempts to identify candidates involved in self-renewal or pluripotency [32–37]. In addition to comparisons with mouse stem cells, a meta-analysis performed by Assou et al., [39] provides further statistical significance to genes that have been found enriched in ESCs by these studies. Importantly, the meta-analysis revealed the ability of the microarray studies to consistently reveal patterns of similarities and differences in gene-expression patterns among human ESC line including high levels of OCT4, SOX2, and NANOG expression in human ESCs, three factors which are critical for maintaining pluripotency in mouse [39–41] and human stem cells [42]. Other genes associated with pluripotency are also found including UTF1, REX1, FOXD3 and members of the FGF2, MAPK-ERK, TGF/activin/nodal, Wnt/b-catenin and Akt/PKB signaling pathways. Together this data demonstrates the efficacy of genomic profiling for determining the “stem-cell” orchestra and provide a foundation for further comparisons of these cells with PGCs and EGCs. Two other studies have also reported on the genomic profiles of PGCs for the purpose of identifying mouse unipotent PGC specific genes that would distinguish them from mouse ESCs [31,43]. Scholer and colleagues [31] revealed 4 new unipotent PGC specific genes that are highly conserved in mouse and human while Mise et al. [43] demonstrated that mouse EGCs were more similar to ESCs compared to mouse PGCs and adult germ line derived stem cells.

In this study, we conduct the first genome-wide comparison study of human PGCs and EGCs with a number of diverse pluripotent stem cell types including ESCs, EGCs, and IPSCs. Genomic analyses classified these cell types into three distinct groups, 1) PGCs, 2) EGCs, and 3) IPSCs. ESCs ECCs, that reflected their developmental potential. These comparisons included qualitative and quantitative differences in known pluripotent associated gene expression as well as revealing a novel signature for human EGCs. Thus, these findings provide important information for identifying potential mechanisms required for PGC reprogramming into the pluripotent state. Comparisons between human PGCs and ESCs will also help facilitate the identification of factors associated with germ line development and may help establish factors related to a common germ cell progenitor potentially shared between these cell types.
Results

Gene expression was studied in five different cell types including PGCs, EGCs, ECCs, IPSCs, and ESCs (Figure 1). In contrast to PGCs, which were directly isolated from tissue, cell lines analyzed in this study were isolated between passage 15 to 27. ECC and ESC lines were XY, while PGCs, EGCs, and IPSCs represented both XX and XY genotypes. PGCs were diploid at this stage in development as previously described [24,44]. Both EGC and IPSC lines were derived from SSEA1+ sorted PGCs and IPSCs generated using lentiviral integration of SOX2, OCT4 and MYCN genes into PGCs. Details of the cells used in this study are described in Table S2 in Supplementary Material.

Quantitative RT-PCR Validation of known Pluripotent and Germ Cell Associated Genes in PGCs, EGCs and Pluripotent Stem Cells

Real-time quantitative RT-PCR analysis (Figure 2) was used to validate the expression of genes known as unique signatures of germ cells and the pluripotent state in our study populations. This was performed on 3 independent cell lines from each group different from those analyzed by microarray analysis. Figure 2 shows data for OCT4, SOX2, NANOG, and DNMT3B from 2–3 independent biological specimens or cell cultures as in the case of ECCs, ESCs, and a human foreskin-derived fibroblast line, HFF1, and each experiment was performed in triplicate. HFFs were used to study relative expression across all groups. Results showed that PGCs expressed higher levels of all pluripotent stem cell genes compared to fibroblasts except DNMT3B. Furthermore, the expression of all pluripotent genes was considerably higher in PGCs than in EGCs. Likewise, EGCs expressed elevated levels of SOX2 similar to ESCs, IPSCs, and ECCs. In fact, this significant increase in SOX2 compared to PGCs in addition to the slightly elevated levels of OCT4 and NANOG expression in EGCs is the most distinguishing feature between these two populations. Although protein levels were not measured in this study, it has been previously reported and shown in our observations that the SOX2 protein is not significantly detectable in human PGCs, unlike mouse PGCs [45]. Therefore, the results of qRT-PCR suggests that the regulation of the SOX2 protein in human PGCs is at the transcriptional level and that regulators of SOX2 gene expression may play a significant role in PGC reprogramming to EGCs. These results are also consistent with microarray and hierarchical analyses which distinguish SOX2 expression as a unique signature of the ESC, IPSC, and ECC group compared to PGCs. Another interesting observation was the down regulation of DNMT3B in both PGCs and EGCs compared to ESCs, IPSCs, and ECCs. This is consistent with the role of DNMT3B in DNA methylation for defining pluripotency in human cells.

Pair-wise Comparisons of Gene Expression Profiles of PGCs and Pluripotent Stem Cell Lines

To explore the similarities in gene expression profiles, pair-wise comparisons and Pearson’s correlations were performed (Table 1). As expected, populations within the same cell types revealed the strongest correlations in gene expression patterns ranging from 0.85–0.99. Variability within groups was the least with IPSCs or ECCs lines which showed the highest correlations within their respective groups at 0.99, P<0.001 each, while PGCs were 0.96 and EGCs ranged from 0.85–0.94. Several possibilities may contribute to the variation found in and among these stem cell lines including the contribution of sex-linked genes and by the passage of time in culture. For instance, when different types of pluripotent stem cell lines are compared, the highest correlations are seen between ECC lines and ESCs (ranging from 0.93–0.94). These lines are both XY and similar in subculture passages. However, EGCs are distinct in their tumorigenic properties from other ESC-like stem cells in that they are malignant cancer stem cells. Nonetheless, the malignant nature of ECCs is not a major contributing factor underlying the close association seen between ESCs. Likewise, it is clear that IPSCs though comprised of both sexes and of an earlier passage than ESCs and ECCs are more similar to male ECCs (0.90–0.91) and ESCs (0.86–0.88) than to either PGCs (0.78–0.83) or ECCs (0.77–0.89). This data is consistent with the pluripotent nature of the IPSCs. Thus, it appears that while sex and subculture passages may contribute to some of the differences seen across cell lines, the close associations found among these cells can be primarily contributed by their pluripotent state.

The expression profiles of EGCs shows the highest similarity to PGCs (~0.89) compared to ECCs, IPSCs, and ESCs (averaging between 0.80–0.82). This is expected as EGCs are derived from PGCs. While the IPSCs and ECCs were also derived by PGCs, it has been hypothesized that ESCs are also derived from an early PGCs progenitor present in the inner cell mass. These results suggest the EGs may represent a distinct state of pluripotency consistent with differences seen between EGCs and other pluripotent stem cells grown under traditional culturing methods. For instance, human EGs, derived under these conditions are more unstable, in that they spontaneously differentiate in culture. EGCs also do not form teratomas in immunocompromised mice even though they can generate a variety of cell types in vitro. Thus, the expression profile of EGCs compared to their unipotent progenitor and other pluripotent stem cells may be indicative of the EGCs unique pluripotent state. Additionally, it would be important to delineate the fraction of genes that may be affected by these cell culturing attributes. Such examination could benefit from the comparisons performed herein among these cell types.

Multivariate Comparisons of Gene Expression Profiles of PGCs and Pluripotent Stem Cells

Global gene expression patterns of PGCs, EGCs, ECCs, IPSCs, and ESCs were analyzed using PCA, which reduces redundancies in variability within high dimensional array data into a smaller number of principal components. Data for PCA including log ratio values and P-values are included in Table S3. Figure 3A shows the plotted position of each cell population against the PC1 and PC3 axes and Figure 3B is representative of the PC1, PC2, and PC3 axes in three dimensional (3D) space. All three PCs accounted for about 71% of the variation present in the entire data set (PC1; 0.445, PC2; .164, PC3; .102). From Figure 3A a distinguishable grouping difference can be seen between ESCs, ECCs, and IPSCs which are grouped together and distinctly located away from EGCs and PGCs along the PC1 and PC3 axis. For instance, PGCs are located in the bottom left side of the PCA plot while EGCs, IPSCs, and ESCs are shifted toward the right along the PC1 component axis and EGCs are grouped in the upper left corner along the PC3 axis. Thus, most of the gene expression differences (44.5%) accounted for were along the PC1 axis and may be to the developmental state of these cells. There was also a greater difference along the PC3 axis between EGCs and PGCs compared to ESCs, ECCs, and IPSCs located in between. Thus, the genes associated with the PC3 axis (constituting 10% of the variation) are likely to define differences in the intermediate pluripotent state of EGs that are independent of the primed human ESC-like state. Figure 3B is a three-dimensional plot representing signature genes of each cell type and correlates.
with Figure 3A in terms of their location. Together, these results identify three distinguishing groups of cells which comprise 1.) unipotent PGCs, 2.) EGCs, and 3.) primed human ESC-like stem cells encompassing ECCs, IPSCs and ESCs.

Next, we tried to detect genes that define the characteristics of the pluripotent stem cell and PGC groups based on the PCA results. Because PGCs are well separated along the PC3 axis, genes that make a large contribution to PC3 were sought using a loading scatter plot, shown as Figure 3C. Grey dots represent
genes which were not in the top differentially expressed genes shown in red. Genes represented by red circles in Figure 3C were the most highly differentially expressed genes (1000 of 54,000, or 1.85% genes in the array) (Table S4) compared to the overall mean across all cell lines with FDR adjusted P<0.0001. The PGC group samples were located between lines at angles of 3.75 and 4.05 radians on the 2D PCA plot (Figure 3A), and genes mapped in the corresponding space by blue lines in the loading scatter plot (Figure 3C) showed elevated expression in the PGC group. These signature genes show enhanced expression in PGCs compared to other cell groups. Genes plotted between 2.10 and 2.80 radians (purple lines in Figure 3C) represent “EGCs signature” genes. Green dotted lines represent genes upregulated in IPSC, ESC and ECC groups located between 5.65 and 0.69 radians.

The location of the specific signature genes along the PC1 and PC3 axis indicates that there is a clear classification or grouping difference among the PGC, EGC, and ESC-like stem cell groups. EGC signature genes are located in the upper half of the PC3 axis and includes genes SRY (sex determining region Y)-box 9 (SOX9),

Table 1. Pair-wise comparisons of gene expression profiles.

<table>
<thead>
<tr>
<th></th>
<th>EGC1</th>
<th>EGC2</th>
<th>EGC3</th>
<th>ESC</th>
<th>PGC1</th>
<th>PGC2</th>
<th>EGC1</th>
<th>EGC2</th>
<th>EGC3</th>
<th>ESC</th>
<th>PGC1</th>
<th>PGC2</th>
<th>IPSC1</th>
<th>IPSC2</th>
<th>IPSC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGC1</td>
<td>0.8997</td>
<td>0.8987</td>
<td>0.8327</td>
<td>0.8931</td>
<td>0.8560</td>
<td>0.8494</td>
<td>0.8469</td>
<td>0.8555</td>
<td>0.8466</td>
<td>0.8698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGC2</td>
<td>0.8469</td>
<td>0.9399</td>
<td>0.8034</td>
<td>0.9174</td>
<td>0.8975</td>
<td>0.7709</td>
<td>0.7671</td>
<td>0.7887</td>
<td>0.7897</td>
<td>0.7669</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGC3</td>
<td>0.8987</td>
<td>0.9399</td>
<td>0.8295</td>
<td>0.9091</td>
<td>0.8858</td>
<td>0.8011</td>
<td>0.7983</td>
<td>0.8076</td>
<td>0.8091</td>
<td>0.7892</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESC</td>
<td>0.8327</td>
<td>0.8034</td>
<td>0.8295</td>
<td>0.8731</td>
<td>0.8372</td>
<td>0.9356</td>
<td>0.9346</td>
<td>0.8685</td>
<td>0.8712</td>
<td>0.8779</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGC1</td>
<td>0.8931</td>
<td>0.9174</td>
<td>0.9091</td>
<td>0.8731</td>
<td>0.9606</td>
<td>0.8363</td>
<td>0.8338</td>
<td>0.8313</td>
<td>0.8349</td>
<td>0.8220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGC2</td>
<td>0.8560</td>
<td>0.8975</td>
<td>0.8858</td>
<td>0.8372</td>
<td>0.9606</td>
<td>0.7983</td>
<td>0.7958</td>
<td>0.7927</td>
<td>0.7962</td>
<td>0.7829</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECC1</td>
<td>0.8494</td>
<td>0.7709</td>
<td>0.8011</td>
<td>0.9356</td>
<td>0.8363</td>
<td>0.7983</td>
<td>0.9949</td>
<td>0.8993</td>
<td>0.8997</td>
<td>0.9133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECC2</td>
<td>0.8469</td>
<td>0.7671</td>
<td>0.7983</td>
<td>0.9346</td>
<td>0.8338</td>
<td>0.7958</td>
<td>0.9949</td>
<td>0.8977</td>
<td>0.8922</td>
<td>0.9138</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSC1</td>
<td>0.8855</td>
<td>0.7887</td>
<td>0.8076</td>
<td>0.8685</td>
<td>0.8313</td>
<td>0.7927</td>
<td>0.8993</td>
<td>0.8977</td>
<td>0.9958</td>
<td>0.9746</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSC2</td>
<td>0.8846</td>
<td>0.7897</td>
<td>0.8091</td>
<td>0.8712</td>
<td>0.8349</td>
<td>0.7962</td>
<td>0.8992</td>
<td>0.9958</td>
<td>0.9974</td>
<td>0.9747</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSC3</td>
<td>0.8698</td>
<td>0.7669</td>
<td>0.7892</td>
<td>0.8779</td>
<td>0.8220</td>
<td>0.7829</td>
<td>0.9133</td>
<td>0.9138</td>
<td>0.9746</td>
<td>0.9747</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0039088.g002

doi:10.1371/journal.pone.0039088.t001
Fibronectin 1 (FN1), FK506 binding protein 6 (FKBP6), Krüppel-like Factor 4 (KLF4) and AXL receptor tyrosine kinase (AXL) which are significantly up-regulated in these cells. These genes may contribute to the partially reprogrammed state of EGCs in culture. PGC signature genes located in the lower half of the PC3 axis include hemoglobin alpha 1 (HBA1), X (inactive)-specific transcript (XIST), testis expressed 13A (TEX13), and SPO11 meiotic protein (SPO11). Spo11, Hba1 and Tex13 have been identified in mouse microarrays as germ cell markers [31]. XIST is known to be downregulated in pluripotent stem cells compared to PGCs [46]. XIST expression was exclusively expressed by the female PGCs as expected and represents the more differentiated state of these unipotent progenitors compared to the female EGC and iPSC which also included.

Genes with shared expression levels in PGCs and EGCs were also identified and may represent a common germ cell progenitor phenotype. For example, located in the middle of the PC3 axis, commonly enriched genes in EGCs and PGCs included imprinted maternally expressed transcript protein (H19), transforming growth factor beta 1 (TGFB1), integrin alpha 8 (ITGA8), insulin-like growth factor binding
protein 5 (IGFBP5), biglycan (BGN), and Moloney leukemia virus 10-like 1 homolog (MOV10L1). Genes that are associated with the PC1 axis define unique differences in pluripotency between PGC, EGC, and the primed human ESC-like groups. This is shown by the distinct localization of known pluripotent stem cell marker genes including AAVOG, MYCN, SOX2, OCT4, and GFP3 in the region containing EGCs, ESCs, and iPSCs. In summary, PCA demonstrates that EGCs exhibit a unique genetic signature from PGCs and other pluripotent stem cells suggesting that EGCs represent a distinctive pluripotent state with many shared features of ESCs/iPSCs and EGCs. Moreover, these results reveal a unique set of genes which may be associated with the pluripotent state of EGCs. Thus, it is possible that these genes are uniquely turned on during the EGC stage of pluripotency and then turned off toward a more naïve like state. Consequently, it would be interesting to compare human EGCs with partially reprogrammed iPSCs and human iPSCs and ESCs to provide further insight into this issue.

Differences in Gene Expression Profiles between PGCs and Pluripotent Stem Cell Lines Define Unique Genetic Signatures of Developmental Potency

Expression analysis of PGCs with the pluripotent stem cell lines indicated that the expression profiles of PGCs and EGCs were significantly different from the ESC, IPSC and ECC groups. The number of differentially expressed genes in EGCs, ESCs, iPSCs, and ECCs compared with PGCs are shown in Table S5. PGCs showed higher expression of 103 genes and significantly lower expression of 656 genes when compared to ESCs (Table S5, Sheet 1 and 2, respectively). Similar numbers of genes were also distinguished when PGC gene expression was compared to either iPSCs or ECCs. For instance, when PGC and ECC profiles were compared 132 genes were upregulated in PGCs compared to the up-regulation of 716 genes in ECCs (Table S5, Sheet 7 and 8, respectively). Similarly, when PGC profiles were compared to iPSCs, PGCs showed higher expression of 233 genes and decreased expression of 496 genes (Table S5, Sheet 5 and 6, respectively). Together, this data shows that a significantly higher number of genes are upregulated in the ESC, IPSC, and ECC lines compared to PGCs. Upregulation of these genes may be relevant to maintaining a primed hESC-like pluripotent state in these lines. In contrast, fewer genes were upregulated in EGCs compared to PGCs. (194 upregulated versus 94 down regulated genes) (Tables S5, Sheet 3 and 4, respectively). Thus, more genes were altered in PGCs during their reprogramming into iPSCs, ESCs, and ECCs than EGCs potentially signifying a unique developmental state for human EGCs.

Specifically, when PGCs are compared to all lines including EGCs, the number of differentially expressed genes is reduced approximately 15 fold. In this case, 20 genes were detected that were upregulated in all four stem cell lines compared to PGCs versus 332 genes in ESCs, iPSCs, and EGCs with EGCs excluded (Tables S5, Sheet 10 and 12, respectively). Likewise, fewer genes are also found upregulated in PGCs when EGC profiles were combined with the analysis, i.e. 49 genes are upregulated in PGCs compared to all pluripotent lines compared to the 70 genes upregulated in PGCs when EGCs are not included (Tables S5, Sheet 9 and 11, respectively). Therefore, the comparisons with EGCs identified a reduced number of candidate genes which may be associated with the conversion of PGCs into the pluripotent state.

Notably, when PGCs were compared to all lines including EGCs, 20 novel associated genes were discovered that were upregulated in the stem cell lines. These included Importin 7 (IPO7), mediator complex subunit 7 (MED7), RNA binding motif protein 26 (RBM26), and heat shock 60 kDa protein 1 (HSPD1). In addition, multiple genes with known or suspected roles in self-renewal or "stemcellness" were highly-up-regulated in ESCs, iPSCs, and ECCs compared to human PGCs. These included SOX2, cyclin E1, (CCNE1), cyclin B1 (CCNB1), and cyclin dependent kinase 6 (CDK6).

Signature Genes Demonstrate Unique Trends in their Expression Pattern Across Lines

Expression patterns of representative genes identified by the PCA as significantly upregulated are shown in Figure 4. PGC signature genes included HBA1, DMR1, TEX13, and meiotic protein SPO11. PGCs do not form tumors while EGCs form malignant carcinomas, iPSCs form benign teratomas, and human EGCs form cyst like structures, not teratomas. Therefore, differential upregulation of expressed genes in ECCs which include sal-like 4 (SALL4), growth differentiation factor 3 (GDF3), v-myc myeloblastosis viral related oncogene, neuroblastoma derived (MYCN), and piwi-like 2 (PIWIL2) may be indicative of their oncogenic as well as their pluripotent properties in these stem cells. Indeed, all four genes have been associated with tumorigenecity, and they demonstrate a similar pattern of expression that was lowest in EGCs and iPSCs compared to ESCs.

In comparison, genes identified in the top 1000 differentially expressed genes transcripts of EGCs included FN1, FKBp6, AXL, and SOX9. FN1, AXL, and SOX9 demonstrated a similar pattern of expression across all lines and have been shown by others to be over expressed in germ cell tumors [47–50]. In contrast, FKBp6 demonstrated a different pattern in which it was uniquely upregulated in EGCs and PGCs but down regulated in ESCs, iPSCs, and ECCs. In mouse studies, this marker was significantly down regulated in both germ line stem cells and ESCs compared to mouse PGCs, and therefore concluded to be a unipotent progenitor marker of mouse PGCs. Although this trend was similar in most of the human cell lines studied here, FKBp6 was also elevated in EGCs. Thus, it is uncertain whether the difference between these two studies is due to speciation or whether up regulation of FKBp6 is uniquely upregulated in EGCs compared to human germ line stem cells.

Interestingly, genes up regulated in iPSCs included both proliferative and anti-proliferative responses controlling cell cycle progression. For instance, iPSCs like other pluripotent stem cells expressed facilitators of the cell cycle such as CCNE1, CCNB1, CDK6 and CDK1. However, iPSCs distinguished themselves from ESCs and ECCs by their elevated expression of known anti-proliferative regulators phosphatidylinositol 3′-kinase receptor 3 (PIK3R3), BCL2-associated X protein (BAX), MAD2 mitotic arrest deficient-like 1 (MAD2L1), and adenomatous polyposis coli (APC). Previous reports have also shown similar results in iPSCs generated from somatic cells with up-regulation of these and other anti-proliferative factors involved in apoptosis, senescence and/or cell-cycle arrest (reviewed in [51]). Similar to these studies, the iPSC profiles here showed up-regulation of the apoptotic mediated factor, BAX, and G1 cell cycle arrest facilitator, P16INK4a (or CDKNA2a). They also expressed elevated levels of PIK3R3 which is a member of the PI3K family and interacts with retinoblastoma protein to regulate cell proliferation and cell cycle progression. As such over expression of PIK3R3 has been associated with ovarian, liver, prostate and breast cancers [52]. When mutated, the tumor suppressor gene, APC, is associated with chromosome instability and tumor progression via beta-catenin signaling [53,54], and when normally expressed is a critical component of cellular defense mechanisms involving cell cycle.
Table 2. Classification of genes differentially expressed in ESCs, EGCs, ECCs, and IPSCs versus PGCs.

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Genes</th>
<th>Explanation</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGC vs. ESC</td>
<td>105</td>
<td>Genes up-regulated in PGCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>656</td>
<td>Genes up-regulated in ESCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td>PGC vs. EGC</td>
<td>194</td>
<td>Genes up-regulated in PGCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>Genes up-regulated in EGCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td>PGC vs. IPSC</td>
<td>233</td>
<td>Genes up-regulated in PGCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>496</td>
<td>Genes up-regulated in IPSCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td>PGC vs. ECC</td>
<td>132</td>
<td>Genes up-regulated in PGCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>716</td>
<td>Genes up-regulated in ECCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td>PGC vs. (ESC+IPSC+ECC)</td>
<td>70</td>
<td>Genes up-regulated in PGCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>332</td>
<td>Genes up-regulated in (ESC+IPSC+ECC)</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td>PGC vs. (ESC+IPSC+ECC)</td>
<td>49</td>
<td>Genes up-regulated in PGCs</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>Genes up-regulated in (ESC+IPSC+ECC)</td>
<td>FC&gt;2, p&lt;0.0001</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0039088.t002

arrest, DNA damage, and repair or by inducing apoptosis [55–57]. Together, these results support evidence in IPSCs derived from somatic cells, which also demonstrate elevated expression of anti-proliferative regulators such as these. Thus, it would be interesting to validate these results and determine whether these factors are involved in the unique signature of IPSCs as a response of their derivation or a necessary requirement of their artificially induced pluripotent state.

Genes up-regulated in ESCs include several interesting targets that have shown some implications of regulating pluripotency in mouse cells. These are shown in Figure 4 which includes deleted in azoospermia-like (*DAZL*), *CYCLIN G1* (*CCNG1*), AT rich interactive domain 2 (*JARID2*), and zinc finger and SCAN domain containing 1 (*ZSCAN1*). *DAZL* has been identified as an early germ cell marker supporting the notion of a germ cell origin for ESCs. In the same family as *ZSCAN1*, *ZSCAN4* has been shown to be critical in regulating telomerase activity and pluripotency of mouse ESCs [30]. *CYCLIN G1* has roles in cell cycle progression and has been shown to be elevated in mouse EGCs and ESCs [59]. More recently, *JARID2* has been shown to be elevated in mouse ESC and modulate pluripotency via polycomb regulation [60].

Comparing Gene Expression Patterns in PGCs, EGCs and Pluripotent Stem Cells using Hierarchical Clustering

To further examine the relationship between human PGCs and EGCs to other pluripotent stem cell lines, signature genes were hierarchically clustered to determine their relationship across all cell lines. Hierarchical analyses was performed for genes that demonstrated the highest differential expression in pluripotent stem cell lines compared to PGCs. Approximately 60% of these genes comprised three distinct regions of gene clusters that were indicative of the PGC, EGC, or the ESC, IPSC and ECC groups. The heat maps of these regions are shown in Figure 5. Consistent with the PCA results, the highest levels of gene expression in PGCs compared to the pluripotent stem cells consisted of a tight cluster of several potential PGC signature genes. These genes corresponded with the loading scatter plot results and are shown to be up-regulated in Figure 5A including *HBB1*, *Doublesex and mab-3 related transcription factor 1 (DMRT1)*, and *Sperm protein associated with the nucleus, X-linked, family member A1 (SPANXA1)*. These predicted signature PGC genes were also identified in the top 1000 differentially expressed genes from PCA. A different pattern between EGCs and PGCs is seen in Figure 5B where many of the genes that are highly up-regulated in EGCs compared to the ESC, IPSC, and ECC group were also elevated in PGCs, though not as high. These included *H19*, *IGFBP5*, *ITGA8*, *BCN* and *EH-domain containing 2 protein (EHD2)*. These results further suggest that EGCs still retain marks of their unipotent PGC-like state. In contrast, *KLF4*, *SOX9*, *AAL*, and *FAN1* were uniquely upregulated in EGCs compared to PGCs suggesting their involvement in EGC derivation. Similar up regulation of KLF4 expression has been demonstrated in the conversion of mouse PGCs into EGCs [31].

Cluster analyses also revealed the overall expression patterns of potential pluripotent genes elevated in the ESCs, ECCs, and IPSCs compared to PGCs and EGCs (Figure 5C). As expected, *SOX2* was significantly elevated in the pluripotent stem cells compared to PGCs consistent with results from PCA. Other factors that were elevated in ESCs, ECCs, IPSCs included those involved in cell cycle namely *CCNB1*, *CCNE1*, *CDK1*, *CDK6*, *HSPA1L*, *v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS)* as well as those defined by the PCA including *PO7*, *MED7*, and *RBM26* (Table S4).

Gene Classification and Biosystems Modeling of Potential Pluripotent Pathways

To identify potential pathways and interpret gene relationships distinguishing EGCs from ESCs, IPSCs and ECCs, a gene classification approach was used by employing the Panther Classification System and Ingenuity Pathways Analysis (IPA).
Figure 4. Identification of potential signature genes of PGCs, ECCs, EGCs, IPSCs, and ESCs. The vertical axis of each graph shows the log ratio of the expression data relative to the population mean. (A) Genes that are up-regulated in PGCs include SPO11, DMRT1, TEX13, and HBA1. (B) Genes up-regulated in ECCs include SALL4, GDF3, MYCN, and PIWIL2. (C) Genes up-regulated in EGCs include FN1, FKBP6, AXL, and SOX9. (D) Genes up-regulated in IPSCs include MAD2L1, PIK3R3, BAX and APC. (E) Genes up-regulated in ESCs include DAZL, CCNG1, JARID2, and ZSCAN1.

doi:10.1371/journal.pone.0039088.g004
Gene classifications were performed using the Panther Classification System, and IPA was used to detect molecular connections between genes that were differentially expressed in pluripotent stem cells compared to PGCs. Together these analyses aid in the understanding of known genes in the context of biological pathways, functions and cellular processes that distinguish EGCs and PGCs from pluripotent stem cells. GO analysis shows the distribution of different cell functions, as percentages, that are allocated in ESCs, IPSCs and ECCs compared to EGCs (Figure 6). Distinct differences between the allocations of biological functions in EGCs compared to these stem cells are found. One of the major differences is their commitment to the cell cycle and metabolic processes.

![Figure 5. Hierarchical clustering of potential signature genes.](image)

Gene expression levels of (A) PGC signature, (B) EGC and common PGC/EGC signature and (C) ESC, IPSC and ECC group genes are represented in a heat map. Lists of genes corresponding to each group are on the right hand side of the cluster tree. Order of the genes in the tables corresponds to their order in the heat map (high expression in red, log_{10} = 1.00; low expression in green; log_{10} = -1.00).

doi:10.1371/journal.pone.0039088.g005
processes while EGCs allot 3.4% to cell cycle and 24.7% to metabolic processes. This is consistent with their reduced proliferation rates in culture. Additionally, there is a distinct difference in the percentage effort allocated for apoptotic functions in EGCs. Specifically, EGCs commit a significantly lower portion, 1.7%, of their cellular functions for regulating apoptosis, compared to ESCs, IPSCs and ECCs which allocate 4.8% percent for that same function. Taken together, these results suggest a strong role for cell cycle and apoptotic factors in maintaining a balance in stem cell self-renewal capacities. Another large difference exists between the commitments of these cells to adhesion processes. EGCs seem to allocate a larger percentage of their cellular demand (9%) to functions relating to cell adhesion while ESCs, IPSCs and ECCs allocate only 1.7%. This is consistent with a primary difficulty in maintaining EGCs in culture which includes their resilience to dispersion during subculture compared to other pluripotent stem cells. Despite the clear differences in allocation of biological processes existing between EGCs and pluripotent stem cells, there are also many similarities. For example, both cells express nearly identical percentages for transport functions, developmental processes, and cellular component organization.

ipa software was also used to detect potential pathways that were specifically up-regulated in pluripotent stem cells and down regulated in PGCs. Figure 7 represents the biological relationships that exist between the regulated proteins responsible for pluripotency. Results showed that as expected the majority of these genes were expressed at similar levels (gray) among pluripotent stem cells and PGCs while SOX2 and FRIZZLED were uniquely up regulated in EGCs, ESCs, IPSCs, and ECCs compared to PGCs. Other known pluripotent regulators, Nanog and oct4, were also shown to be expressed but at similar levels among groups. Most interestingly, when genes upregulated in the pluripotent stem cells compared to PGCs was analyzed, IPA analysis identified an integral network that appears to be upregulated in pluripotent stem cells. Figure 8 is a graphical representation of intersecting networks responsible for controlling cell cycle, DNA replication, DNA repair, recombination, and cell death in which the majority of these components were upregulated in the ESC, IPSC and ECC group compared to PGCs. This included cyclins Ccne1, Ccng1 and Ccnb1 as well as apoptotic and DNA repair-recombination regulators, Bax and Poly (ADP-ribose) polymerase 1 (Parp1).
When EGCs were included in the comparisons, two genes with ties in cell cycle regulation, *KRAS* and *HSPD1*, were upregulated in all stem cells compared to PGCs. *KRAS* is a GTPase in the *Ras* family and is essential in normal tissue signaling of PI3-kinase but elevated in many cancers where it suppresses tumor suppressor genes. More recently, *KRAS* has also been associated with the undifferentiated state of pancreatic cancer stem cells [61] and in testicular germ cell tumors [62]. Like *KRAS*, *HSPD1* plays active roles in cell signaling processes and has been associated with tumorigenesis. This heat shock protein is also associated with regeneration in lower vertebrates and recently shown to be controlled by two known regulators of pluripotency, LIN28 and LET-7, (for review [63]) during retinal regeneration in zebrafish [64]. Proteomic analyses have also detected elevated expression of *HSPD1* in mouse and human ESCs [65–69].

**Discussion**

This study is the first to report a comparative analysis of human EGCs and PGCs with other pluripotent human stem cells. These results demonstrate a unique pattern of expression for human EGCs that is distinct from PGCs and other pluripotent stem cells. Differences are also seen between human PGCs and EGCs with their mouse counterparts. Specifically, this study identified 20 novel genes that were upregulated in all pluripotent stem cells compared to PGCs and 78 genes upregulated specifically in EGCs. These novel genes provide evidence for unique factors in germ cells that may contribute to their developmental state and potentially to the intermediate pluripotent state of EGCs. These genes comprised a significant portion of regulators of the cell cycle, DNA repair, and DNA recombination. This was supported by GO analysis which demonstrated a substantial amount of energy exerted in cell cycle and metabolic processes in the stem cell lines.
compared to PGCs and by elevated utilization of these processes in the ESCs, IPSCs, and ECCs compared to EGCs.

Unique Features of the PGC and EGC Transcriptional Program

Despite being committed to a single lineage, PGCs are unique because they co-express many key pluripotent genes. Specifically, we show here for the first time that human PGCs and EGCs show quantitative differences in their expression of pluripotency associated genes, SOX2 and DNMT3B, compared to other human pluripotent stem cells. SOX2, together with OCT4 and NANOG, is required for ESC cell self-renewal and pluripotency (reviewed in [70,71]) and for reprogramming somatic cells into IPSC cells [72]. Others have similarly reported that SOX2 protein is not expressed in human PGCs [45]. Specifically, our data shows that human PGCs express SOX2 transcript albeit at reduced levels (similar to fibroblast cell lines) while it is significantly elevated in human EGCs, IPSCs, ECCs and ESCs.

In contrast, DNMT3B levels in PGCs and EGCs were similar to fibroblasts compared to elevated levels in the ESC, IPSC, and ECCs.
ECC lines. DNMT3B levels are critical for DNA methylation and its expression elevated in human ESCs [34,37]. DNMT3B has also been identified as one of three genes critical to distinguishing the fully reprogrammed state of pluripotent human ESCs and iPSCs [74]. Here, we show for the first time considerably lower expression of this gene in human PGCs and EGCs compared to levels detected in ESCs, iPSCs, and ECCs further suggesting that human EGCs exist in a unique pluripotent state and that reduced DNMT3B expression may be a contributing factor for the unique chromatin state of EGCs [75].

In addition to known markers of pluripotency, the analyses performed here reveal unique factors of the germ cell lineage that may also define the developmental potency of human PGCs and EGCs. Specifically, this study found several genes that were highly expressed in either human PGCs exclusively or also in EGCs compared to other pluripotent stem cells suggesting their role in regulating the unique states of these cell types. Genes upregulated in PGCs alone included HBA1, SPACA1, and DMRT1 while genes upregulated in both PGCs and EGCs included EHD2, ITGA8, epithelial membrane protein 1 (EMPI), and collagen, type I, alpha 2 (COL1A2). These genes are interesting candidates for future studies as other reports have either directly or indirectly associated their elevated expression with the germ cell lineage yet none have reported their role in regulating germ line developmental potency. For instance, DMRT1, up regulated in human PGCs, is a tumor suppressor gene with putative roles in regulating PGC proliferation. The absence of Dmrt1 causes mice spermatogonia to precociously exit the spermatogonial program and enter meiosis [76]. It has also been shown to directly repress Sox2 expression in mice and when down regulated increase teratoma formation. Reduced expression has also been identified in human germ cell tumors [77]. Thus, DMRT1 is a potential candidate for repressing SOX2 expression in human PGCs. Likewise, EHD2, ITGA8 and EMPI have also been identified as tumor suppressor genes that when altered is associated with highly malignant ovarian cancers and germ cell tumors [62,78-81]. Likewise, we show elevated expression of COL1A2 in human PGCs and EGCs similar to others who have shown that COL1A2 distinguishes type A spermatogonia stem cells from differentiated germ cells in mouse [82].

Thus, genes found upregulated exclusively in PGCs in this study are potential candidates involved in maintaining the unique state of pre-meiotic, proliferative human PGCs while those upregulated in PGCs and EGCs may play a role in maintaining a unique pluripotent state of human EGCs.

Other genes that were among the most highly upregulated in PGCs included 3 of 11 genes recently identified that distinguished mouse PGCs from pluripotent germline stem cells. These genes included germ cell specific proteins like TEX13 and meiotic protein SP011, as well as the hemoglobin protein, HBA1 [31]. Three other markers that were discussed in the report including, Pik3r3, Mov10l1, and Fkbp6, were also differentially expressed in our human cell populations. However, unlike mouse PGCs, where these genes were upregulated, in human PGCs PHR3 and FKBP6 expression was consistently down regulated compared to human stem cells, and MOV1BL1 was expressed at similar levels between human PGCs and EGCs. While these results suggest specie differences in their expression in human and mouse cells, future confirmation is warranted.

Unique Features of the Stem Cell Transcriptional Program

From the top most differentially expressed genes, several candidates that may play a role in controlling pluripotency were discovered. These genes were highly up regulated in EGCs, iPSCs, ECCs and ESCs compared to PGCs and included genes involved in chromatin remodeling such as IP07, MED7, and RBM26 as well as DNA repair and transcriptional activation such as HSPD1 and KRAS. These candidates are particularly interesting as recent evidence has shown roles for these factors in mechanisms that facilitate pluripotency. For instance, one report has suggested an important role of IP07 in the nuclear import of Sox2 and high mobility group (Hmg) box domain proteins in mouse ESCs [83] which are two well-known regulators of stem cell self-renewal. Another interesting candidate that is upregulated in EGCs compared to PGCs was MED7. Mediator proteins, like MED7, are known traditionally for their role as transcriptional co-activators required for RNA polymerase II activity [72]. However, mutations in MED7 have recently identified a novel role for MED7 in directly silencing subtelomeric genes and increasing telomere length and life span in Saccharomyces cerevisiae. This makes MED7 an interesting candidate given that telomere activity is also known to play a significant role in regulating pluripotency.

KRAS and HSPD1 play active roles in cell signaling processes. Here we show for the first time their potential involvement in establishing the pluripotent state of human stem cells as both genes were among a few genes that were highly differentially expressed in all stem cell lines compared to PGCs. The heat shock protein, HSPD1, is known as a regeneration-associated gene in lower vertebrates and recently shown to be controlled by two known regulators of pluripotency, LIN28 and LET-7 [64], in zebrafish [74]. Likewise, KRAS has also been shown to be repressed by LET-7 in cancer cell lines resulting in reduced radio sensitivity [84]. This role of KRAS would make it consistent with a molecule that is involved in contributing to the pluripotent stem cell phenotype as stem cells are well known for their increased sensitivity to DNA damage.

It is known that pluripotent stem cells have a distinct cell-cycle from differentiated cells. They exhibit long-term proliferative capacity by spending a proportionally shorter period of time in G1 and a proportionally longer period of time in S phase compared to adult cells [85,86]. However, it is still unknown how these distinguishing features contribute to the pluripotent state. Here, we report for the first time differences in expression of cell cycle components between human PGCs and pluripotent germline stem cells which may implicate their role in defining the pluripotent state. These included both G1 and G2 phase mediators such as CCNB1, CCNG1, MAD2L1, Aurora kinase A (AURKA), HSPD1, and KRAS. For instance, CCNE1 and CDK6 along with G1 checkpoint mediator MAD2L1 were upregulated in ESCs, iPSCs, and ECCs but not EGCs compared to PGCs. While a few studies have shown the role of CYCLIN D and CYCLIN E in mediating the stem cell phenotype of ESCs their role in maintaining the pluripotent state of EGCs has not been studied [87-90]. Likewise, a similar pattern of expression in constituents of the G2/M phase including CYCLIN B, CYCLIN G and CDK1 along with G2/M checkpoint regulator AURKA was also revealed to be upregulated in iPSCs, EGCs and ESCs compared to EGCs. Thus, it remains unknown how differences in G1 and G2 cell cycle contribute to the unique pluripotent state of EGCs which appear more similar to PGCs in their cell cycle expression than other pluripotent stem cells.

Links between pluripotency and self-renewal have been established with examples of pluripotency associated genes modulating key regulators of cell cycle and vice versa. For example, SOX2/S and OCT4 (O) through microRNAs, and NANOG (N) has been shown to regulate G1 progression at the transcription level [84]. Quantitative RT-PCR analysis performed in this study specifically showed that human EGCs expressed intermediate levels of these pluripotent genes and that SOX2 was
one of the top down regulated genes in PGCs. Therefore, discovering potential links between these three genes and how they regulate the cell cycle may unravel key mechanisms involved in germ cell reprogramming. Interestingly, two cell cycle regulators potentially involved in the SON network MYCN and CYCLIN E which have established roles in defining pluripotency in human stem cells and which were down regulated in EGCs compared to other human stem cell lines [91] are likely targets for contributing to the EGC’s unique pluripotent state.

Important to the self-renewal of pluripotent stem cells is the sensitivity of these cells to DNA damage. As stem cells give rise to all cell types of the embryo, mutations incurred in early development would be detrimental to an organism [11,92–96]. Thus, stem cells are hypersensitive to DNA damage [97] and as a result, demonstrate more effective stress defense pathways than more differentiated cell types. Examples of these mechanisms include removing endogenous free radicals generated by their increased proliferation [97,98] and greater efficiency in DNA damage repair after radiation compared to their differentiated counterparts [99]. Thus, pluripotent stem cells must elicit unique responses to DNA repair and recombination. Our study reports for the first time novel connections with key regulators of DNA repair and recombination in regulating the pluripotent state including KRAS and PARP1. When activated, KRAS suppresses DNA repair-related tumor suppressor genes involved in homologous recombination (HRR) such as BRCA1, BRCA2, TP53, and EXO1 [100]. Specifically, it has recently been shown that HRR is the predominant DNA double-strand break repair pathway in mouse ESCs with minimal contributions by nonhomologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) repair mechanisms [101]. This is consistent with NHEJ and MMEJ being more error-prone (reviewed in ref. [102–105]) and theoretically resulting in unsustainable mutations. In a similar fashion, elevated expression of PARP1 in stem cells suggests that it may also be associated with pluripotency. For instance, Poly (ADP-ribose) polymerase proteins, like PARP1, are involved in a number of cellular processes involving mainly in DNA repair and programmed cell death and comparisons performed here support data that have shown PARP1 is involved in the OCT4 and SOX2 network in mouse ESCs [24,44].

Conclusion

In conclusion, this study is the first to compare and analyze global gene expression profiles of human EGGS and PGCs with those of embryo-derived and germ-cell derived stem cells. The results reveal distinct features regarding the transcriptional programs of these cell types. Results from this study identified sets of genes that characterize the developmental status of human EGGS and PGCs from ECCs, IPSCs, and ESCs, and serve as important genes to delineate germ cell lineages and pluripotency in human cells. Furthermore, these findings provide important data for identifying potential mechanisms required for the reprogramming of unipotent primordial germ cells into the pluripotent state. Comparisons of PGCs identified 20 genes that are specifically upregulated in all stem cell types including EGGS provides important information for distinguishing differential pluripotent states in human germ cells and possibly conventional ESCs and IPSCs as well.

Summary

In summary, differences in the transcriptional profiles and signature genes across different stem cell types shown here are consistent with multiple states of the pluripotency. In particular, EGGS exhibit a genetic signature distinct from other pluripotent stem cells which suggests that these stem cells are in a unique pluripotent state. These factors include novel regulators of the cell cycle, DNA repair and recombination. However, it remains to be seen whether EGGS are more like naive mouse ESCs or whether they represent a partially reprogrammed IPSC-like state. Alternatively, differences in their genetic profiles would attribute to a notion that not all pluripotency-associated genes are regulated in the same way in all pluripotent stem cells, or that epistatic interactions play a significant role in the defining pathways that generate the undifferentiated state.

Materials and Methods

Collection of Tissue

Gonadal tissues, primordial germ cells and embryonic germ cell lines were obtained from human fetuses 8–11 weeks post fertilization as a result of termination of pregnancy, via protocols and written patient consent, approved by the Joint Committee on Clinical Investigation of the Johns Hopkins University School of Medicine. Gestational age was confirmed by anatomical markers which include limb and digit formation, crown heel and crown rump measurement as well as the first day of the last menstrual cycle. Ages are discussed in terms of fetal development and not the age from the last menstrual period. Sex of the tissue was determined by gross morphological examination of the gonads and by fluorescent in situ hybridization of tissue connected to the gonads as previously performed [7,24].

Human PGC Acquisition and EGC Derivation

PGCs were isolated using magnetic cell sorting technology (MACs) and indirect labeling of cells with magnetically tagged goat anti-mouse IgM antibodies toward a mouse-anti-SSEA1 antibody (Miltenyi Biotech). Briefly, gonads were minced in 1 mg/mL collagenase, incubated at 37°C for 20 min, rinsed, and incubated with SSEA1 antibody (1:5 dilution) for 15 min on ice. Afterward, secondary antibody was applied at 1:100 dilution for another 30 min on ice and sorted on magnetic columns as previously described [24,106,107]. SSEA1+ PGCs were either directly prepared for microarray analyses or used to generate EGGS. For EGC generation, SSEA1+ PGCs from a single gonad were sorted and approximately 50 cells were seeded in each of 12 wells of a 96-well plate with irradiated mouse embryonic feeder cells, SIM 6-thioguanine resistant ouabain (STO) (~125,000 cells/well; ATCC) [106]. Media consisted of Dulbecco’s modified Eagle’s medium-199 (Invitrogen) supplemented with 20% Knockout serum (Invitrogen), 2 ng/mL FGF2 (R&D Systems), 1000 U LIF (Millipore), 10 mM forskolin, and 20 ng/mL BMP4 (R&D Systems).

Human ESC Culture

Human ESCs from the H1 line (WiCell, Fed ID # 0043) were cultured on matrigel (BD Biosciences) in 10-cm cell culture dishes with Dulbecco’s modified Eagle’s medium/F12-Knockout serum-based media conditioned by mouse embryonic fibroblast cells (MEFs) (Millipore, Strain CF1) and supplemented with 4 ng/mL FGF2 as described previously [106].

Human ECC Culture

The human embryonal carcinoma line NTERA-2 cl.D1 and Tera-2 was acquired through American Type Culture Collection (ATCC) (Virginia) and cultured on matrigel-coated plates under conditions described previously for this cell line [109].
Human Fibroblasts

HFF-1 line was acquired by ATCC (SCRC-1041) and is a human fibroblast cell line originally derived from the foreskins of two individuals [110]. Cells were grown in similar condition as MEFs (Millipore, Strain CF1) in Dulbecco’s Modified Eagle’s Medium, DMEM199 supplemented with 15% bovine serum albumin.

IPSC Generation and Culture

All IPSC lines were derived from PGCs obtained by different specimens. Lentiviral production were performed using Gateway®-compatible cDNAs for SOX2, OCT4, and MYCN (Invitrogen) inserted via recombination with LR Clonase into a modified (attR1, attR2) pLVX-Puro vector (Clontech) using the Gateway® recombination system (Invitrogen). Purified, high titer VSV-G pseu typed lentiviral preps for expression of each transgene were produced using standard methods. SSEAl+ PGCs were transduced with replication-defective recombinant lentivirus for 12 hours at multiplicity of infections (MOI) of 5–10 for each construct in the presence of 6 ug/ml Polybrene (Sigma).

Lentiviral expression was confirmed by qRT-PCR analysis with transgene-specific primers. Transgenes were silenced three weeks after lentiviral transfections. Pluripotency was confirmed by transgene-specific primers. Transgenes were silenced three weeks at multiplicity of infections (MOI) of 5–10 for each construct produced with replication-defective recombinant lentivirus for 12 hours at multiplicity of infections (MOI) of 5–10 for each construct in the presence of 6 ug/ml Polybrene (Sigma).

Micro-array Analysis

To analyze gene expression profiles, the Affymetrix Human U133 Plus 2.0 GeneChip was used. RNA extraction, reverse transcription, cRNA preparation, and chip hybridization were performed according to the manufacturer’s instructions (Affymetrix, Santa Clara, CA). In brief, total RNA was extracted from cultured cells using RNeasy (Qiagen, Valencia, CA) protocols described below for RT-PCR. Five micrograms of purified RNA were then used as a template for double-stranded cDNA synthesis primed using a T7-c(dT)24 oligonucleotide. Double-stranded cDNA was then used as a template for biotin-labeled cRNA preparation using T7 RNA polymerase. Biotinylated cRNA (15 µg) was fragmented at 94°C for 35 minutes (100 mM Tris-acetate pH 8.2, 500 mM potassium acetate, 150 mM magnesium-acetate), and hybridized to the Affymetrix HG U133 Plus 2.0 GeneChips containing ~54,000 transcripts for 16 hours at 45°C with constant rotation (60 rpm). An Affymetrix Fluidics Station 450 was used to remove the non-hybridized target and to incubate with a streptavidin-phycocerythin conjugate to stain the biotinylated cRNA. The staining was amplified using goat IgG as blocking reagent and biotinylated goat anti-streptavidin antibody, followed by a second staining step with a streptavidin-phycocerythin conjugate. Fluorescence was detected using the Affymetrix G3000 GeneArray Scanner and image analysis of each GeneChip was done through the GeneChip Operating System 1.4 (GCOS) software from Affymetrix, using the standard default settings. Statistical analyses of microarray data were performed using a combination of bioconductor (http://www.bioconductor.org) and Partek™ software (Version 6.5) (http://www.partek.com). The raw signal values were normalized with quantile normalization method and gene level expression was summarized using RMA (Robust Multi-Array) method [111]. Principal Component Analysis (PCA) were performed using Partek™ software and differential gene expression were detected using bioconductor package limma [112]. The FDR adjusted p value cutoff of <0.0001 was used to obtain the lists of differentially expressed genes. Hierarchical clustering analysis was also performed using Spotfire™.

Quantitative Real-Time RT-PCR

RNA from EGCs, PGCs, ESCs, ECCs, and IPSCs were isolated for quantitative real-time (qRT)-polymerase chain reaction (PCR) using MiniRNeasy kits (Qiagen 74124) with the RNA clean-up protocol and optional on-column DNase treatment. Complementary DNA was generated with SuperScript III First-Strand Synthesis System RT Kits, following the manufacturer’s instructions (INV18080-051). Real-time qRT-PCR analysis was performed using ABI7900HT with Taqman Assay-on-Demand designed oligonucleotides for the detection of OCT4, SOX2, NANOG and DNM3T3 and each sample had a template equivalent to 3 ng of total RNA (Table S1). Quantitation was measured using the DDCt method and normalized to b-actin. Each primer set was tested in at least triplicate across biological replicates.

Statistics

t-Tests were performed to evaluate the significance between two groups. Significance was accepted at p<.05.

Gene Classification and Biosystem Modeling

The Ingenuity Pathway Analysis (IPA) program (http://www.ingenuity.com) was used for pathway and gene classification analysis of differentially expressed genes. The microarray data set was translated into HUGO gene identifiers and uploaded to the IPA system. The IPA software is a Java based online exploratory tool with a curated database for genes with millions of published literature references. The IPA database builds gene networks, pathways, and biological function clusters. IPA software uses published literature from the database to map the biological relationship of the uploaded genes. Fisher’s exact test is used to determine the probability that each biological function is due to chance alone. Scores for IPA networks are the negative logarithm of the p-values calculated. They indicate the likelihood of the focus proteins in a network being found together due to random chance. Scores of 2 or higher have at least a 99% likelihood of not being generated by chance alone. Panther Classification System was used to perform gene ontology analyses. This system uses the Gene Ontology (GO) platform to classify genes by biological process, molecular function and cell components and includes commonly used classes of protein functions many of which are not covered by other GO analyses (www.pantherdb.org).

Data Access

Supplemental material will be provided at www.genome.org. Raw microarray CEL files will be deposited in the GEO database with accession numbers upon acceptance of the publication.

Supporting Information

Table S1 The real-time qRT-PCR analysis data for the detection of OCT4, SOX2, NANOG and DNM3T3B.

Table S2 Details of the cells used in this study.

Table S3 Data for PCA results including log ratio values and P-values.

Table S4 The most highly differentially expressed genes (1000 of 54,000 or 1.85% genes in the array)
represented by the red circles in Figure 3C. These genes are compared to the overall mean across all cell lines with FDR adjusted P<0.0001.

Table S5 Complete list of differentially expressed genes in EGCs, ESCs, IPSCs, and ECCs compared to PGCs.

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>

Acknowledgments

The authors would like to thank Cynthia F. Liu for her contribution to this work.

Author Contributions

Conceived and designed the experiments: CK. Performed the experiments: CK NP. Analyzed the data: NP CK SG HH AA RC ADLA JG. Contributed reagents/materials/analysis tools: CK AA. Wrote the paper: NP CK SG.


